View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

VALUE IN HEALTH 16 (2013) 670–685

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Computer Modeling of Diabetes and Its Complications: A Report on the Fifth Mount Hood Challenge Meeting

Andrew J. Palmer, BmedSci, MBBS, The Mount Hood 5 Modeling Group*

Menzies Research Institute Tasmania, University of Tasmania, Hobart, TAS, Australia

ABSTRACT

Objectives: The Mount Hood Challenge meetings provide a forum for computer modelers of diabetes to discuss and compare models, to assess predictions against data from clinical trials and other studies, and to identify key future developments in the field. This article reports the proceedings of the Fifth Mount Hood Challenge in 2010. **Methods:** Eight modeling groups participated. Each group was given four modeling challenges to perform (in type 2 diabetes): to simulate a trial of a lipid-lowering intervention (The Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus [ASPEN]), to simulate a trial of a blood glucose–lowering intervention (Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation [ADVANCE]), to simulate a trial of a blood pressure–lowering intervention (Cardiovascular Risk in Diabetes [ACCORD]), and (optional) to simulate a second trial of blood glucose–lowering therapy (ACCORD). Model outcomes for each challenge were compared with the published findings of the respective trials. **Results:** The results of the models varied from each other and, in some cases, from the published trial data in important ways. In general, the models performed well in terms of predicting the relative benefit of interventions, but performed less well in terms of quantifying the absolute risk of complications in patients with type 2 diabetes. Methodological challenges were highlighted including matching trial end-point definitions, the importance of assumptions concerning the progression of risk factors over time, and accurately matching the patient characteristics from each trial. **Conclusions:** The Fifth Mount Hood Challenge allowed modelers, through systematic comparison and validation exercises, to identify important differences between models, address key methodological challenges, and discuss avenues of research to improve future diabetes models.

Keywords: computer simulation, cost-effectiveness analysis, diabetes, health economics, modeling.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

A decade after Jonathan Brown and Andrew Palmer met to compare the 20-year predictions of two computer simulation models of type 2 diabetes in the Timberline Lodge, high on the side of Mount Hood near Portland, Oregon, the fifth Mount Hood Challenge meeting was held in Malmö, Sweden, in September 2010 [1]. A total of eight modeling groups participated in the 2010 challenge, which followed a similar format to previous Mount Hood meetings whereby modelers were asked to use their prediction models to simulate the outcomes of clinical studies to inform debate on the challenges facing groups working in this area.

Computer simulation models, in essence a series of mathematical equations combined in a structured framework, have many uses such as allowing data from clinical trials to be extrapolated over longer time periods and to other populations. By providing information for health care decision makers on long-term clinical outcomes and costs, such models allow informed choices to be made between available interventions. As the issue of cost containment becomes ever more pertinent for many health care decision makers, the reliance on computer simulation modeling is increasing. This is particularly true of chronic diseases such as type 2 diabetes, which develop over a long period of time and are associated with significant morbidity and mortality and a substantial economic burden [2].

Although cost-of-illness studies have taught us a great deal about the scale of the economic burden associated with diabetes, as well as the identity of the main cost drivers, they do little to help us understand the incremental value of new interventions in a given population. Clinical trials provide essential information on new interventions, but their limitations in terms of time frame (typically 1–3 years), tightly controlled designs, and often (highly) selected populations can make their findings difficult to generalize to other care settings or populations. Key parameters such as demographics, life expectancy, patient management/medical technology, treatment costs, and health budgets can vary widely between regions and between countries. Flexible computer models have the potential to overcome these problems and provide valuable information, such as assessments of long-term cost-effectiveness, for policymakers and reimbursement decision makers. To fulfill this role, models must be based on

* *Address correspondence to:* Andrew J. Palmer, Menzies Research Institute Tasmania, Medical Science 1 Building, 17 Liverpool Street (Private Bag 23), Hobart, TAS 7000, Australia.

E-mail: andrew.palmer@utas.edu.au.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.01.002

the best available evidence, and validated against clinical data (internal and external validation) as well as each other, and they must also be transparent, documented in detail, and open about their mechanisms and assumptions.

The aim of this article was to report the proceedings of the Fifth Mount Hood Challenge held in Malmö, Sweden, in 2010, with a view to providing a summary of how eight current diabetes models match up to data from published clinical studies as well as to each other, to highlight differences between models, and to offer an insight into the challenges facing diabetes models a decade after the First Mount Hood Challenge.

Research Design and Methods

The Fifth Mount Hood Challenge was the first meeting for 6 years (since the Fourth Mount Hood Challenge in 2004 [3]), and participating modelers were asked to perform simulations based on four published clinical trial data sets, thereby allowing comparison of all eight participating models against clinical data. Treatments and interventions, management of patients, and cohort characteristics were defined in advance to minimize the number of potentially disparate assumptions required to make reliable forecasts over the duration of follow-up reported in the trial publications. The working hypothesis for the Mount Hood Challenge was that this process of standardized comparison is the best method to identify differences between models as well as assess the models' reliability in predicting the consequences of changes in risk factors brought about by an intervention in a clinical trial situation. Readers are referred to the Mount Hood web site for details of instructions given to modeling groups, and for contact details of modeling groups for further clarifications or detailed information regarding individual model structures and assumptions [4]. All groups with a published simulation model of type 2 diabetes were invited to take part in the challenge. In total, eight groups accepted this year's challenge. They were joined at the meeting by 85 participants from 10 countries.

Simulation Challenges in Type 2 Diabetes

To expand on the validation exercises from the Fourth Mount Hood Challenge Meeting in 2004, the modelers performed four external validation analyses against three recent clinical trials that reported the results of interventions attempting to modify key risk factors for the complications of type 2 diabetes. For each of these trials, the modeling groups attempted to predict the event rates of the primary end points and of as many secondary outcomes as possible at the end of the study by using only the size of the risk factor change achieved where possible within the individual models' frameworks, or using relative effects of treatments if this was the only option with a model. All modeling groups were restricted from using any information that they may have had access to (e.g., patient-level data) that was not in the public domain to produce the primary results set for presentation at the meeting. Such data, however, could be used to produce an additional set of results to examine whether model performance could be improved. The end points reported (by default) by the models did not always identically match those reported in the clinical trials. While modeling groups were asked to report as many outcomes as they could for each trial, some models could not generate results for every end point. As a result, the results tables are incomplete, with empty cells where no appropriate end point for comparison was available from that model. Notable differences in end-point definitions have been cited where relevant in this article. Results were to be reported as the proportion of patients experiencing each type of event (so as to match the outcomes reported in the trial publications). The following challenges were set.

Lipid-lowering intervention based on the Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus trial [5]

The Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus (ASPEN) was a 4-year, double-blind, parallel group trial of 10 mg of atorvastatin versus placebo in patients with type 2 diabetes and low-density lipoprotein cholesterol levels below contemporary guideline targets ($\leq$160 mg/dl [4.1 mmol/l], or $\leq$140 mg/dl [3.6 mmol/l] for subjects with myocardial infarction [MI] or coronary intervention >3 months before screening). The composite primary end point comprised cardiovascular death, nonfatal MI, nonfatal stroke, recanalization, coronary artery bypass surgery, resuscitated cardiac arrest, and worsening or unstable angina requiring hospitalization. Exclusion criteria included hemoglobin A_{1c} (Hb A_{1c}) value of over 10% (86 mmol/mol), blood pressure over 160/100 mm Hg, body mass index (BMI) over 35 kg/m^2, preexisting liver disease, kidney disease, or heart failure treated with digoxin. Patients were advised to adopt a National Cholesterol Education Program diet (which is low in saturated and *trans* fats, and rich in fruits, vegetables, whole grains, fat-free and low-fat dairy products, and lean meat, fish, and poultry). A total of 2410 patients were randomly allocated to receive atorvastatin or placebo. Mean patient age was 60 years, and approximately two-thirds of the study population was male (Table 1). At the end of the study, composite primary end point rates for atorvastatin and placebo were 13.7% and 15.0%, respectively (hazard ratio 0.90; 95% confidence interval [CI] 0.73–1.12). Subgroup analysis in patients with a history of MI or interventional procedure showed a hazard ratio of 0.82 (95% CI 0.59 –1.15). In patients with no prior MI or interventional procedure, the hazard ratio was estimated to be 0.97 (95% CI 0.74–1.28) for atorvastatin versus placebo.

Blood glucose–lowering intervention from the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial [6]

In the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) trial, a total of 11,140 patients with type 2 diabetes were randomly assigned to undergo either standard glucose control or intensive glucose control, defined as the use of gliclazide (30–120 mg modified release) plus other drugs (metformin, thiazolidinediones, acarbose, and/or insulin) as required to achieve an Hb A_{1c} value of 6.5% or less (Table 1). Primary end points were composites of major macrovascular events (death from cardiovascular causes, nonfatal MI, or nonfatal stroke) and major microvascular events (new or worsening nephropathy or retinopathy), assessed both jointly and separately. Inclusion criteria included a history of macro- and microvascular disease and age 55 years or more. Exclusion criteria included a requirement for insulin at the time of study initiation. After a median 5 years of follow-up, mean Hb A_{1c} value was lower in the intensive-control group (6.5%) than in the standard-control group (7.3%). Intensive control reduced the incidence of combined major macrovascular and microvascular events (18.1%, vs. 20.0% with standard control; hazard ratio 0.90; $P = 0.01$), as well as that of major microvascular events (9.4% vs. 10.9%; $P = 0.01$), primarily because of a reduction in the incidence of nephropathy. The hazard ratio for macrovascular events was 0.94 for intensive treatment versus standard care ($P = 0.32$, not significant).

Table 1 – Summary of population characteristics in the four challenges.

Characteristic	ASPEN	ADVANCE	ACCORD BP	ACCORD BG
Number of patients	2,410	11,140	4,733	10,251
Mean age (y)	60.0	66.3	63.2	62.2
Percentage males (%)	66	57	53	39
Percentage current smokers (%)	12	15	13	14
Mean body mass index (kg/m^2)	29.0	27.0	32.2	32.2
Mean Hb A_{1c} value (%)	8.0	8.0	8.0	8.0
Mean fasting plasma glucose (mmol/l)	–	8.5	8.4	9.74
Median duration of diabetes (y)	8.0	8.2	9.8	10.0
Mean systolic blood pressure (mm Hg)	133.0	145.4	147.0	131.0
Mean diastolic blood pressure (mm Hg)	76.0	80.7	80.4	74.8
Mean LDL-cholesterol (mmol/l)	2.94	3.11	2.85	2.69
Mean HDL-cholesterol (mmol/l)	1.22	1.20	–	–
Mean HDL-cholesterol (mmol/l) female/male	–/–	–/–	1.33/1.08	1.22/1.00
History of cardiovascular disease (%)	–	–	32	30
History of myocardial infarction (%)	17	12	–	–
History of cerebrovascular disease or stroke	5	9	–	–

ACCORD, Action to Control Cardiovascular Risk in Diabetes trial; ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation; ASPEN, Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus; BP, blood pressure; BG, blood glucose; Hb A_{1c}, hemoglobin A_{1c}; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

Blood pressure–lowering intervention from the Cardiovascular Risk in Diabetes trial [7]: The Action to Control

Cardiovascular Risk in Diabetes (ACCORD) trial was designed to investigate whether therapy targeting normal systolic blood pressure (SBP) reduces major cardiovascular events in participants with type 2 diabetes at high risk for cardiovascular events. A total of 4733 participants with type 2 diabetes were randomly assigned to intensive therapy (target SBP <120 mm Hg) or standard therapy (target SBP <140 mm Hg). The primary composite end point was nonfatal MI, nonfatal stroke, or death from cardiovascular causes. Mean follow-up was 4.7 years. Inclusion criteria included age 40 years or more, atherosclerosis, left ventricular hypertrophy, or two or more risk factors from a list of serum lipid levels, smoking, hypertension, and obesity. Patients were excluded if they had an Hb A_{1c} value below 7.5% (58 mmol/mol) or BMI above 45 kg/m^2. Baseline cohort characteristics are summarized in Table 1. The annual rate of the primary outcome was 1.87% in the intensive-therapy group and 2.09% in the standard-therapy group (hazard ratio 0.88; $P = 0.20$, not significant). The annual rates of death from any cause were 1.28% and 1.19% in the two groups, respectively ($P = 0.55$, not significant), but the annual rates of stroke, a prespecified secondary outcome, were significantly different (0.32% vs. 0.53%, $P = 0.01$).

Blood glucose–lowering intervention from the ACCORD trial [8]

The ACCORD trial also examined the efficacy of intensive glucose lowering–therapy on mortality and key cardiovascular events in patients with advanced type 2 diabetes and a high risk of cardiovascular events. In the glucose-lowering study, 10,251 patients with a median Hb A_{1c} value of 8.1% were assigned to receive intensive therapy (target Hb A_{1c} value below 6.0%) or standard therapy (targeting an Hb A_{1c} value of 7.0%–7.9%). Of these patients, 38% were women and 35% had had a previous cardiovascular event (Table 1). The primary outcome was a composite of nonfatal MI, nonfatal stroke, or death from cardiovascular causes (i.e., the same composite as used in the blood pressure study). Because of a higher mortality rate in the intensive-therapy group, intensive therapy was discontinued after a mean of 3.5 years of follow-up. At the end of follow-up, the primary outcome occurred in 352 patients (6.9%) in the intensive-therapy group, versus 371 (7.2%) in the standard-therapy group (hazard ratio 0.90; $P = 0.16$, nonsignificant). Two hundred fifty-seven patients (5%) in the intensive-therapy group died, however, compared with 203 patients (4%) in the standard-therapy group (hazard ratio 1.22; $P = 0.04$).

Overview of Models Participating in the Fifth Mount Hood Challenge

Eight models were represented at the meeting (in order of presentation): the IMS CORE Diabetes Model, the University of Michigan Model for Diabetes (Michigan Model), The Swedish Institute of Health Economics model entitled the Economics and Health Outcomes in Type 2 Diabetes Mellitus (ECHO-T2DM) Model, the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model, The UKPDS Risk Engine, the Centers for Disease Control (CDC)-RTI Diabetes Cost-effectiveness Model, the Cardiff Research Consortium Model (Cardiff Model), and the Evidence-Based Medicine Integrator (EBMI) Simulator. The models presented results on all four validation challenges based on the outcomes reported by each respective model. The EBMI Simulator did not report results on the ASPEN challenge. Each presenting model team prepared a brief model description, which is provided in the following paragraphs (in order of presentation).

IMS CORE Diabetes Model [9,10]

The IMS CORE Diabetes Model is a non–product-specific, diabetes policy analysis tool that performs real-time simulations. Disease progression is based on a series of interdependent Markov submodels that simulate diabetes-related complications (angina, MI, congestive heart failure, stroke, peripheral vascular disease, diabetic retinopathy, macular edema, cataract, hypoglycemia, ketoacidosis, nephropathy and end-stage renal disease, neuropathy, foot ulcer, and amputation). Each submodel uses time-, state- and diabetes-type–dependent probabilities derived from published sources, and utilizes tracker variables to overcome the memoryless properties of standard Markov models. The progression of relevant physiological parameters (e.g., Hb A_{1c}, SBP, lipids, and BMI) is simulated on the basis of long-term epidemiological data, and event risk is constantly updated on the basis of these risk factors. Analyses, including first- and second-order Monte

Carlo simulations, can be performed on patient cohorts with either type 1 or type 2 diabetes, defined in terms of age, gender, baseline risk factors, preexisting complications, and comorbidities. The model is adaptable, allowing the inclusion of new clinical and economic data as these become available. The creation of country-, health maintenance organization-, or provider-specific versions of the model is possible. The reliability of simulated clinical outcomes has been tested, with results validated against those reported from clinical trials and epidemiological studies.

The Michigan Model [11,12]

The Michigan Model for Diabetes has been substantially revised since its original publication in 2005 and is implemented by using newly developed software that models chronic diseases. This software provides an environment for model design, estimation, and simulation, as well as a convenient graphical user interface to 1) define parameters, 2) define populations, 3) generate populations from distributions, 4) create a new disease model or modify an existing model, 5) simulate the behavior of a given base population by using a defined model enhanced by a set of simulation rules, and 6) analyze and report simulation results. These capabilities support models expressed as multiple nested extended Markov subprocesses. It uses Monte Carlo simulation to simulate disease progression. As a result, the software system is very general and can accommodate many chronic disease models in the same software environment, which enables it to compare and contrast results from alternative models.

The current version of the software simulates disease progression in the following disease processes: diabetes, nephropathy, neuropathy, retinopathy, cardiovascular disease (CVD), and cerebrovascular disease. It calculates costs and utility scores for user-specified time periods. Users can specify the frequency of examination (which results in diagnosis) and the rate of compliance with treatment (which affects the rate of disease progression) among other things. The Michigan Model for Diabetes is publicly available under a General Public License and can be downloaded from the Michigan Diabetes Research and Training Center web site (www.med.umich.edu/mdrtc/cores/DiseaseModel/index.html).

ECHO-T2DM [13]

ECHO-T2DM is a (second-order) stochastic, microsimulation model that consists of Markov health states representing the development and consequences of key micro- and macrovascular complications. A cohort of hypothetical type 2 diabetes patients is generated from a probability distribution of initial patient characteristics (both demographic and health-related). Hb A_{1c} is the core driver of the model, affecting both outcomes and changes in treatment. Patients are initially treated with one of two user-specified treatment paradigms, and their evolving health and treatment needs are simulated annually until the end of the user-defined time horizon or death (if occurring sooner). Patient health is recorded by using health states that capture the existence and severity of retinopathy, nephropathy, neuropathy, and CVD (as defined in the UKPDS) and is updated on an annual basis.

Diabetes treatment is governed by an algorithm that seeks to maintain user-defined Hb A_{1c} treatment thresholds. User-defined inputs control the algorithm (making it flexible). Treatment affects Hb A_{1c} but can also affect BMI, SBP, and lipid levels, and can be initiated or discontinued and new agents added to meet Hb A_{1c} goals. Treatments can also cause adverse events, including hypoglycemia, which can lead to discontinuation or noncompliance. The model estimates costs on the basis of the treatment of diabetes-related complications and the costs of blood glucose control. Utility scores are calculated by using decrements for each complication. Model outcomes include incidence rates for each of the complications and adverse events modeled, incremental cost-effectiveness ratios, net monetary benefit results, and acceptability curves.

UKPDS Outcomes Model [14]

The UKPDS Outcomes Model is a widely used simulation model for type 2 diabetes based on patient-level data from the UKPDS. It models the occurrence of seven diabetes-related end points (MI, angina, stroke, heart failure, amputation, renal failure, and blindness in one eye) and death to estimate quality-adjusted life expectancy, life expectancy, and costs. In brief, the UKPDS Outcomes Model is based on an integrated system of parametric equations that predict the annual probability of any of the above end points and Monte Carlo methods to predict the occurrence of events. The likelihood of the end points is based on patient demographics, duration of diabetes, risk factor levels, and history of diabetes-related complications. Different treatment and management strategies are evaluated through their impact on risk factor levels. A key aspect of the model is its ability to capture the clustering or interaction of different types of complications at the individual patient level. The model is a probabilistic discrete-time illness-death model, rather than a Markov model. Patients start with a given health status (e.g., age, sex, duration of diabetes, risk factor values, and no complications) and can have one or more nonfatal complications and/or die in any model cycle by comparing estimated probabilities with random numbers drawn from a uniform distribution ranging from 0 to 1 to determine whether an event occurs. When a patient experiences a complication, their utility is permanently decremented such that they accumulate quality-adjusted life-years at a slower rate. Utility decrements and costs associated with events are estimated from the same patient-level data set. Elements of the UKPDS Outcomes Model have been widely used in many other diabetes simulation models. A software implementation of the model is available under license from the University of Oxford (http://www.dtu.ox.ac.uk/Outcomesmodel).

UKPDS Risk Engine [15–17]

The UKPDS Risk Engine is a type 2 diabetes–specific risk calculator based on over 50,000 patient-years of data collected between 1977 and 2007. It is designed to provide estimates and 95% CIs for CVD risk in individuals with type 2 diabetes not known to have heart disease. Risk estimates can be calculated for any given duration of diabetes on the basis of current age, sex, ethnicity, smoking status, presence or absence of atrial fibrillation, presence or absence of microalbuminuria or worse, and levels of Hb A_{1c}, SBP, total cholesterol, and high-density lipoprotein cholesterol.

The CDC-RTI Diabetes Cost-effectiveness Model [18–20]

The CDC-RTI Diabetes Cost-effectiveness Model is a Markov simulation model of disease progression and cost-effectiveness for type 2 diabetes that follows patients from diagnosis to either death or an age of 95 years. The model simulates the development of diabetes-related complications on three microvascular disease pathways (nephropathy, neuropathy, and retinopathy) and two macrovascular disease paths (coronary heart disease and stroke). In the model, progression between disease states is governed by transition probabilities that depend on risk factors and duration of diabetes. Interventions affect the transition probabilities and resulting complications. Model outcomes include disease complications, deaths, costs, and quality-adjusted life-years. A detailed description of the model and its validation analysis can be found online [17,18].

The Cardiff Model [21–23]

The Cardiff cost-utility model estimates the long-term economics and health impact of managing patients with type 2 diabetes. The core model is coded in C++ and linked to a Microsoft Excel front end. The model is a fixed time increment (yearly) stochastic simulation with a 40-year time horizon. The model utilizes the UKPDS Outcomes Model equations (UKPDS 68 [12]) to predict macrovascular and microvascular complications. The dynamic profile of modifiable risk factors is controlled via user-controllable equations, which also includes dynamic changes to weight. The model is designed to evaluate a treatment and control pathway, each of which is composed of up to three lines of therapy. Therapy escalation is controlled via user-defined Hb A_{1c} thresholds.

The model incorporates the risk of hypoglycemia, which is captured by using annual rates for the occurrence of symptomatic and nocturnal hypoglycemic events; these events are associated with immediate cost and utility consequences. The model is capable of running with mean values, with probabilistic inputs and user-defined data with outputs for cost per life-year gained and cost per quality-adjusted life-year gained, in addition to the number, cost, and utility consequence of all events occurring within the model.

Evidence-Based Medicine Integrator Simulator [24]

EBMI is a free, open-source medical computational framework that integrates three kinds of knowledge by using stochastic discrete-event microsimulation: risk estimates derived from patient data, comparative effectiveness estimates obtained from randomized clinical trials, and genetic knowledge from basic research. EBMI is designed to be used with electronic medical records to identify all potential treatments for every patient and to recommend next treatment steps. It can also generate or accept population data to do management, trial planning, policy, and research studies. EBMI's direct use of local data and clinical trial results maximizes clinical safety. Assumptions are readily changeable as new trials appear. The EBMI code and documentation are available at http://code.google.com/p/ebmi.

Currently, EBMI simulates all major macrovascular and microvascular complications of type 2 diabetes, plus associated expenditures and utility effects. Highly detailed protocols use natural dosage increments for all classes of diabetes and CVD treatments, plus user-defined classes. Event drive protocols are also programmable. EBMI is designed to safely integrate reliable knowledge with local circumstances, not to predict the results of treatments that have not been trialed (although the user can hypothesize effects). Therefore, the traditional hypothesis that simulators should not be validated against trials does not apply to EBMI in the usual way. It is important to demonstrate that EBMI can reproduce the relative effects of treatment trials and also reproduce local data where it is to be used. The EBMI version used in the Fifth Mount Hood Challenge is derived from, and has been validated against, the Kaiser Permanente Northwest diabetes registry.

Results

Model Predictions Based on the ASPEN Clinical Trial (Lipid-Lowering Intervention)

The first of the validations at the Fifth Mount Hood Challenge challenged the diabetes modelers to reproduce the findings of a lipid-lowering therapy from the ASPEN clinical trial in patients with elevated low-density lipoprotein levels. Comparison with the primary composite end point from ASPEN (cardiovascular death, nonfatal MI, nonfatal stroke, recanalization, coronary artery bypass surgery, resuscitated cardiac arrest, and worsening or unstable angina requiring hospitalization) showed that the four models reporting a value underestimated the overall risk in both the atorvastatin and placebo treatment groups (Table 2; Figure 1). Interestingly, the models tended to overestimate the risk of cardiovascular mortality and MI events (except the Cardiff Model, which underestimated the incidence of these events), but underestimated the risk of stroke and notably overestimated the risk of noncardiovascular mortality. This latter observation may suggest that the data on mortality from other causes used in the models do not reflect the standard of care that patients in the ASPEN trial were receiving. Other interesting observations were that some model appeared to perform better for primary versus secondary cohorts (e.g., ECHO and CDC) and that, despite being built using different approaches, the ECHO and CDC models generated comparable results.

In terms of the relative risk of events on atorvastatin versus placebo, which would be important in estimating absolute benefits in a cost-effectiveness analysis where incremental values are reported, the models tended to overestimate the benefit of atorvastatin on the primary end point (Table 2). This was also true of primary cardiovascular events and cardiovascular mortality, although the UKPDS Outcomes Model was an exception in that it reported a relative risk of cardiovascular mortality closely matching that in the ASPEN study. There were no obvious modeling trends in terms of the relative risk of other end points in the ASPEN study from the data presented.

Model Predictions Based on the ADVANCE Clinical Trial (Glucose-Lowering Intervention)

In general, the results produced by the models in the validation against the ADVANCE trial showed a reasonably close match with reported trial results after 5 years of follow-up (Tables 3 and 4; Figures 2 and 3). Models tended to overestimate the incidence of major macrovascular events, with the UKPDS Outcomes Model, the EBMI Simulator, and the IMS CORE Diabetes Model offering the closest estimates to the trial data. Models also tended to overestimate the risk of coronary events (major and all) in both the intensive and standard glucose control treatment groups in the ADVANCE trial. In terms of specific macrovascular events, the models overestimated MI risk, underestimated stroke and heart failure risk, and overestimated CVD mortality. Model estimations of overall mortality (death from any cause) were, as in the validation against the ASPEN study, underestimated by two of the three models reporting this end point.

Estimates of the incidence of microvascular complications showed that models generally underestimated the overall risk of nephropathy and retinopathy as defined in the ADVANCE trial. Similarly, the models reporting values tended to underestimate the risk of new-onset microalbuminuria and new or worsening neuropathy in comparison with the trial data. (This latter end point was challenging for several of the models.)

In terms of the relative risk of complications between the intensive and standard glucose control treatment arms in the ADVANCE trial, the models performed reasonably well in estimating the relative risk of major macrovascular events, major coronary events, nonfatal MI, and heart failure (Tables 3 and 4). In general, the models overestimated the benefit of intensive glucose control on stroke risk and very slightly underestimated the benefit on cardiovascular mortality. In contrast, there was a tendency to underestimate the benefit of intensive glucose control on all-cause mortality. None of the models closely predicted the occurrence of more major cerebrovascular events in the intensive glucose control group of the ADVANCE trial relative to the standard-control group. Overall, the relative risk

Table 2 – Summary of validation results for a lipid-lowering intervention in the ASPEN trial.

Model		Primary composite end point	CVD mortality (Prim/Sec)	Non-CVD mortality (Prim/Sec)	Fatal/nonfatal MI (Prim/Sec)	Fatal/ nonfatal stroke	Angina	Primary event	Secondary event
ASPEN									
Control		15.0%	3.1% (2.5/5.6)	2.6% (2.3/3.6)	5.5% (3.6/12.6)	3.1%	3.2%	10.8%	30.8%
Intervention		13.7%	3.1% (2.0/7.1)	2.6% (2.1/4.8)	4.0% (2.9/8.3)	3.0%	2.8%	10.4%	26.2%
IMS CORE									
Control			3.9% (2.2/10.7)	5.7% (5.1/8.5)	5.8% (4.8/11.4)	2.5%			
Intervention			3.4% (1.8/10.3)	5.5% (4.9/8.4)	4.5% (3.6/9.1)	2.0%			
MICHIGAN									
Control		3.3%		2.3% (2.5/–)	3.9% (3.2/–)		3.4%	2.7%	
Intervention		2.7%		2.8% (2.5/–)	6.1% (5.5/–)		3.0%	2.1%	
ECHO-T2DM									
Control		14.8%	5.0% (4.0/8.0)	5.1% (4.0/11.2)	8.5% (7.1/14.9)			12.7%	23.9%
Intervention		12.3%	4.3% (3.3/7.4)	4.9% (3.9/10.3)	6.7% (5.6/11.9)			10.5%	20.4%
UKPDS OM									
Control		11.1%	4.1%	7.4%	6.1%	2.8%	1.9%		
Intervention		9.6%	4.1%	7.0%	5.2%	2.5%	2.3%		
UKPDS RE									
Control								8.7%	
Intervention								7.4%	
CDC-RTI									
Control		14.3%	5.5% (3.3/13.9)	6.4% (6.3/6.6)	9.4% (8.0/14.4)			12.2%	22.4%
Intervention		12.4%	5.0% (2.8/13.7)	6.4% (6.4/6.6)	8.1% (6.4/14.4)			10.0%	21.5%
CARDIFF									
Control			(1.7/3.1)	(4.1/6.0)	(3.1/5.8)				
Intervention			(1.5/3.5)	(4.1/6.0)	(2.7/5.8)				
ASPEN 2006	Relative risk	0.91	1.00	1.00	0.73	0.97	0.88	0.96	0.85
IMS CORE	Relative risk		0.88	0.97	0.77	0.77			
MICHIGAN		0.81			0.64			0.79	
ECHO-T2DM		0.83	0.87	0.96	0.79			0.82	0.85
UKPDS OM		0.86	1.00	0.95	0.85	0.91	0.84		
UKPDS RE								0.84	
CDC-RTI		0.87	0.91	1	0.86			0.82	0.96
CARDIFF									

Notes. Relative risk represents the ratio of event risk in the intervention group versus the control group. Results were reported in the trials and by some of the modeling groups for all patients, primary prevention patients, and secondary prevention patients. Results in parentheses are for the primary/secondary prevention subgroups, respectively.ASPEN, Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus, a 4-year, double-blind, parallel group trial of 10 mg of atorvastatin versus placebo in patients with type 2 diabetes and LDL-cholesterol levels below contemporary guideline targets; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; LDL, low-density cholesterol; MI, myocardial infarction; MICHIGAN, The Michigan Model for Diabetes; Prim, primary prevention group; Sec, secondary prevention group; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

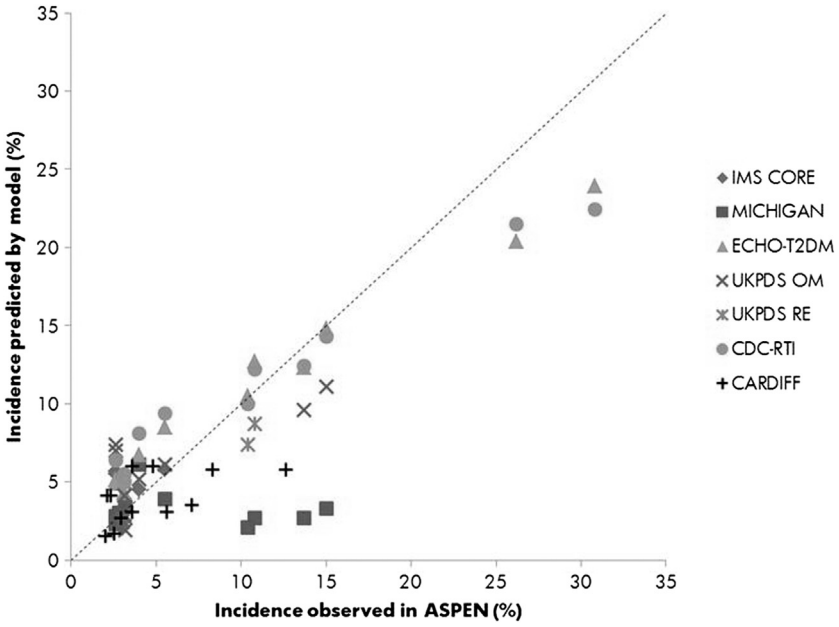

Fig. 1 – Summary of model validation results for the ASPEN trial. ASPEN, Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in Non-Insulin-Dependent Diabetes Mellitus, a 4-year, double-blind, parallel group trial of 10 mg of atorvastatin versus placebo in patients with type 2 diabetes and LDL-cholesterol levels below contemporary guideline targets; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; LDL, low-density lipoprotein; MI, myocardial infarction; MICHIGAN, The Michigan Model for Diabetes; Prim, primary; Sec, secondary; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

of microvascular complications was generally in line with the observations made in the trial, but there was a notable spread of relative risk values for nephropathy and retinopathy between models. The Michigan Model produced some outliers in the validation against the ADVANCE data because of differing endpoint definitions for microvascular complications between the trial and the modeling analysis.

Model Predictions Based on the ACCORD Clinical Trial (Blood Pressure–Lowering Intervention)

In the validation focused on a blood pressure–lowering intervention in the ACCORD trial, the models generally performed well in terms of reproducing the primary end point (nonfatal MI, nonfatal stroke, or death from cardiovascular causes) over 4.7 years of follow-up (Table 5; Figure 4). Most models produced values comparable to the trial data set for the end points of stroke (any) and nonfatal MI, although the IMS CORE Diabetes Model and the Cardiff Model underestimated the risk of MI in both the intervention and control treatment groups. The models, in general, provided overestimates of the risk of all-cause mortality with the exception of the Cardiff Model, which closely matched the trial data. Four models overestimated cardiovascular mortality (the Michigan Model, the ECHO-T2DM Model, the CDC/RTI Model, and the UKPDS Outcomes Model), and two models provided values comparable with the ACCORD data (the IMS CORE Diabetes Model and the Cardiff Model). The IMS CORE Diabetes Model, the ECHO-T2DM Model, and the UKPDS Outcomes Model provided estimates of heart failure risk in line with those observed in the trial, whereas the Cardiff Model and the EBMI Simulator underestimated and overestimated these risks, respectively.

In terms of relative risk, the models performed reasonably well on the primary end point, major coronary events and nonfatal MI, but underestimated the benefit of the blood pressure–lowering intervention on stroke risk (Table 5). For all-cause mortality, none of the models predicted a higher mortality rate in the intervention arm (aggressive blood pressure control) than in the control arm as observed in the trial. Similarly, most models failed to predict a similar trend in cardiovascular mortality in trial data, with the exceptions of the UKPDS Outcomes Model and the Cardiff Model.

Model Predictions Based on the ACCORD Clinical Trial (Glucose-Lowering Intervention)

Validation against the ACCORD blood glucose intervention trial was an exploratory and optional analysis at the Fifth Mount Hood Challenge Meeting (Table 6; Figure 5). The ACCORD study showed that the use of intensive therapy to target normal Hb A_{1c} levels for 3.5 years increased mortality and did not significantly reduce major cardiovascular events, a risk not previously identified in clinical trials in type 2 diabetes [8]. Only the UKPDS Outcomes Model and the ECHO-T2DM Model reported validation results for the primary end point in the ACCORD trial (nonfatal MI, nonfatal stroke, or death from cardiovascular causes; Table 6). The UKPDS Outcomes Model provided estimates close to those reported in the trial and the ECHO-T2DM Model slightly overestimated risk in both the intensive glucose therapy arm and in the control arm. Both models overestimated the benefit of intensive therapy versus control on the primary end point. On the incidence of nonfatal MI, the models offered a range of values that spanned those reported in the trial, with the EBMI Simulator reporting values closest to the ACCORD published data. In general, estimates of nonfatal stroke incidence were comparable with those reported in the ACCORD trial, but all models failed to predict the increased risk of stroke observed in the intensive-treatment arm

Table 3 – Summary of validation results for a glucose-lowering intervention in the ADVANCE trial (primary end point).

Model		CVD mortality	All major macrovascular/ microvascular events	Major macrovascular events	Nonfatal MI	Nonfatal stroke	Major microvascular events	Nephropathy	Retinopathy
ADVANCE									
Intensive		4.5%	18.1%	10.0%	2.7%	3.8%	9.4%	4.1%	6.0%
Standard		5.2%	20.0%	10.6%	2.8%	3.8%	10.9%	5.2%	6.3%
IMS CORE									
Intensive		4.2%	15.7%	11.3%	4.6%	2.6%	4.4%	2.3%	2.1%
Standard		4.6%	17.5%	12.2%	4.9%	2.7%	5.4%	2.8%	2.6%
MICHIGAN									
Intensive		5.6%	56.1%	16.4%	7.4%	4.7%	46.6%	26.3%	27.8%
Standard		5.7%	55.9%	16.7%	7.7%	4.7%	46.0%	25.8%	27.4%
ECHO-T2DM									
Intensive		6.6%	24.1%	16.1%	7.2%	3.0%		2.7%	6.9%
Standard		7.5%	30.1%	18.0%	7.9%	3.5%		4.3%	10.8%
UKPDS OM									
Intensive		6.4%		10.5%					
Standard		6.5%		11.4%					
UKPDS RE									
Intensive				13.1%					
Standard				14.5%					
CDC-RTI									
Intensive		11.0%	26.0%	17.8%			8.2%	2.4%	5.8%
Standard		11.4%	29.5%	19.2%			10.3%	3.2%	7.1%
CARDIFF									
Intensive		2.2%			2.2%	1.6%			
Standard		2.4%			2.1%	1.6%			
EBMI									
Intensive				11.2%*					21%
Standard				12.0%*					21%
ADVANCE	Relative risk	0.87	0.91	0.94	0.96	1.00	0.86	0.79	0.95
IMS CORE	Relative risk	0.92	0.90	0.93	0.94	0.93	0.82	0.82	0.81
MICHIGAN		0.98	1.00	0.98	0.96	0.99	1.01	1.02	1.01
ECHO-T2DM		0.89	0.80	0.90	0.91	0.87		0.64	0.64
UKPDS OM		0.98		0.92					
UKPDS RE				0.91					
CDC-RTI		0.96	0.88	0.93			0.80	0.74	0.83
CARDIFF		0.91			1.03	0.97			
EBMI				0.93					0.99

Note. Relative risk represents the ratio of event risk in the intervention group versus the control group. ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial, a 5-year trial in type 2 diabetes in which patients were randomly assigned to undergo either standard glucose control or intensive glucose control, defined as the use of gliclazide plus other drugs as required to achieve an Hb A_{1c} value of 6.5% or less; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; MICHIGAN, the Michigan Model for Diabetes; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

* Combined fatal and nonfatal MI and stroke.

Table 4 – Summary of validation results for a glucose-lowering intervention in the ADVANCE trial (secondary end points).

Model		Death from any cause	All CVD events	All coronary events	Major coronary events	All cerebrovascular events	Major cerebrovascular events	Heart failure	PVD	New-onset MA	New/worsening neuropathy
ADVANCE											
Intensive		22.1%	8.9%	10.1%	5.6%	6.3%	4.3%	3.9%	6.2%	23.7%	42.4%
Standard		22.4%	9.6%	10.3%	6.1%	5.9%	4.4%	4.1%	6.6%	25.7%	41.5%
IMS CORE											
Intensive		14.9%	13.1%	12.4%	8.7%	2.9%	2.9%	3.3%	1.9%	5.4%	12.6%
Standard		15.9%	13.4%	13.2%	9.4%	3.1%	3.1%	3.4%	2.5%	6.6%	15.4%
MICHIGAN											
Intensive			10.3%							15.9%	22.2%
Standard			10.1%							16.7%	22.2%
ECHO-T2DM											
Intensive		19.4%	16.3%		11.3%	3.6%		2.7%	10.1%	1.4%	
Standard		21.5%	17.3%		12.6%	4.2%		3.0%	10.1%	2.0%	
UKPDS			17.3%		7.2%		3.3%	2.9%			
OM			17.8%		7.7%		3.7%	3.2%			
Intensive											
Standard											
CDC-RTI											
Intensive			16.4%								
Standard			16.7%								
CARDIFF											
Intensive								1.0%			
Standard								1.2%			
EBMI											
Intensive		27.7%	24.4%		4.8%		8.8%	9.1%	2.7%	8.6%	9.4%
Standard		28.7%	25.2%		5.6%		8.8%	9.2%	2.7%	9.6%	9.7%
ADVANCE	Relative risk	0.99	0.93	0.98	0.92	1.07	0.98	0.95	0.94	0.92	1.02
IMS CORE	Relative risk	0.94	0.98	0.94	0.93	0.93	0.93	0.95	0.75	0.81	0.82
MICHIGAN			1.02							0.95	1.00
ECHO-T2DM		0.90	0.94		0.89	0.86		0.91	1.00	0.68	
UKPDS OM			0.97		0.93		0.90	0.90			
CDC-RTI			0.98								
CARDIFF								0.89			
EBMI		0.96	0.97		0.86		1.00	0.99	0.98	0.90	0.97

Notes. The UKPDS Risk Engine did not report appropriate outcomes for the secondary end point analysis of the ADVANCE trial. Relative risk represents the ratio of event risk in the intervention group versus the control group. ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial, a 5-year trial in type 2 diabetes in which patients were randomly assigned to undergo either standard glucose control or intensive glucose control, defined as the use of gliclazide plus other drugs as required to achieve an Hb A_{1c} value of 6.5% or less; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; MA, microalbuminuria; MICHIGAN, the Michigan Model for Diabetes; PVD, peripheral vascular disease; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model.

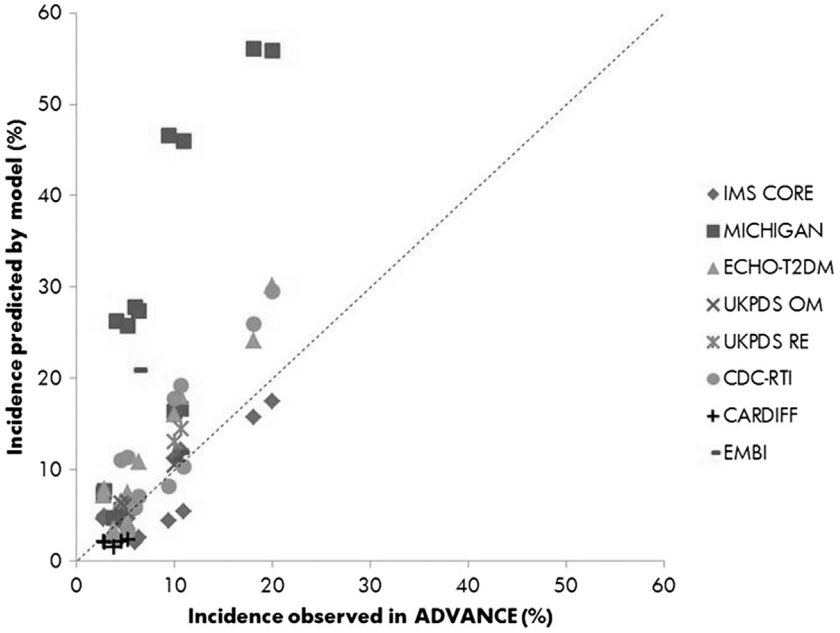

Fig. 2 – Summary of model validation results for the ADVANCE trial (primary end point). ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial, a 5-year trial in type 2 diabetes in which patients were randomly assigned to undergo either standard glucose control or intensive glucose control, defined as the use of gliclazide plus other drugs as required to achieve an Hb A_{1c} value of 6.5% or less; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; MICHIGAN, the Michigan Model for Diabetes; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

of the ACCORD trial. The models tended to overestimate the incidence of cardiovascular mortality in the ACCORD population, the exceptions being the IMS CORE Diabetes Model and the Cardiff Model, which reported values closest to those observed in the trial. However, all models failed to predict the increased risk of cardiovascular mortality associated with the intensive intervention in the ACCORD trial. Modeled estimates of all-cause mortality were generally higher than those reported in the ACCORD trial. The exception here was the Cardiff Model, which reported all-cause mortality incidences close to those in the ACCORD trial. Only the Cardiff Model and the EBMI Simulator predicted an increased risk of all-cause mortality with intensive therapy in the ACCORD population, although the relative risks reported were not a close match for the value from the ACCORD data set.

Discussion

The general consensus on the results presented at the Fifth Mount Hood Challenge was that, in general, the models performed reasonably well in terms of predicting the relative risk of interventions versus control treatments, but less well in terms of the estimation of absolute risk. While relative effects are important to many decision makers, predictions of absolute effects may be needed, not just the relative changes if absolute amounts of money spent or saved need to be predicted, or how many adverse clinical events will be prevented or caused. Absolute rates of events may also be important to people designing clinical trials, as they directly affect the power of the trial and therefore the sample size, duration, and cost. When discussing the difficulty of simulating the trials, it is important to distinguish between simulating the outcomes in the control group versus the outcomes in the intervention group. For some of the trials (e.g., ACCORD glucose lowering) modeling the effect of the intervention was challenging, whereas the outcome rates in the control groups can still provide very useful tests of the models' accuracies in addressing risk factors and the progression of complications with current care. The selection of trials for the validation exercise at the Fifth Mount Hood Challenge was deliberately challenging. For example, the ASPEN study was specifically chosen because it produced a different result (i.e., no statistically significant benefit of atorvastatin treatment on the primary composite end point) from Collaborative Atorvastatin Diabetes Study (CARDS), which was used in a previous Mount Hood Challenge in 2004. The ADVANCE study reported data from a fairly atypical type 2 diabetes population: patients were generally old, with long-standing advanced diabetes but not on insulin, and with over one-third of the population recruited to study centers in Asia. Historically, there has been a paucity of data available to inform the modeling of "high-risk" populations such as those in ADVANCE and ACCORD. Moreover, the relative lack of data on the cardiovascular risk profile in Asian patients with type 2 diabetes (relative to their Western counterparts) has made accurate modeling of outcomes challenging in the ADVANCE population. Data from the ACCORD trial has challenged previously accepted wisdom on the role of aggressive treatment of risk factors in diabetes. Aggressive treatment of blood pressure did not lead to a statistically significant composite outcome of fatal and nonfatal major cardiovascular events in the ACCORD trial, and intensive therapy targeting Hb A_{1c} value below 6% was shown to increase mortality without significantly reducing major

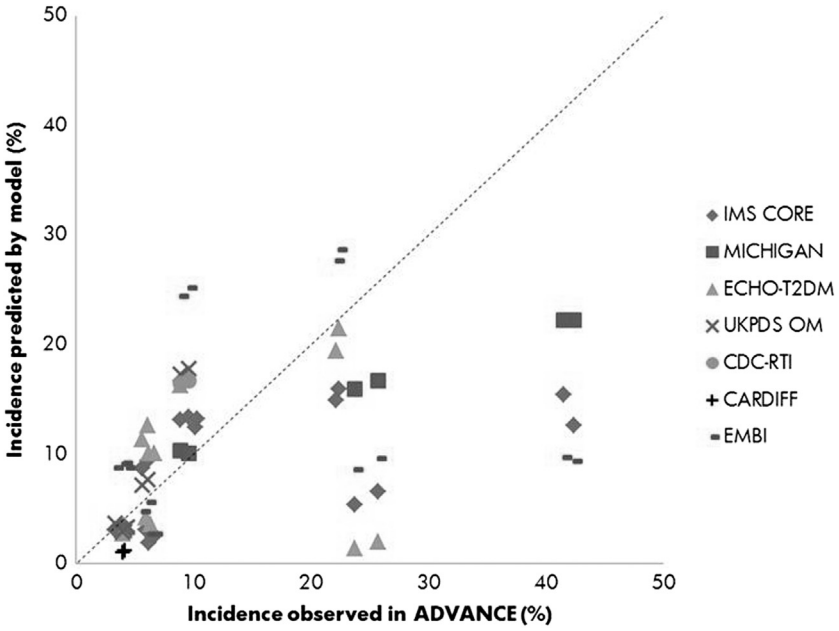

Fig. 3 – Summary of model validation results for the ADVANCE trial (secondary end points). The UKPDS Risk Engine did not report appropriate outcomes for the secondary end point analysis of the ADVANCE trial. ADVANCE, Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation trial, a 5-year trial in type 2 diabetes in which patients were randomly assigned to undergo either standard glucose control or intensive glucose control, defined as the use of gliclazide plus other drugs as required to achieve an Hb A_{1c} value of 6.5% or less; CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model ; IMS CORE, IMS CORE Diabetes Model; MA, microalbuminuria; MICHIGAN, the Michigan Model for Diabetes; PVD, peripheral vascular disease; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model.

cardiovascular events. Because many of the models rely primarily on data linking risk factors to hard end points from landmark studies that predate these more recent and perhaps atypical data sets, the validation exercise set for the Fifth Mount Hood Challenge was a very demanding one. Many of the models were not able to reproduce all the primary outcomes and many of the secondary outcomes of each of the trials.

During the meeting discussion session, the modeling groups raised a number of issues that may have contributed to the discrepancies in absolute risk between the model predictions and the trial results. Matching population characteristics were cited by several groups as a testing aspect of the validation analyses (due to complicated inclusion and exclusion criteria in the trials). Difficulties around effectively modeling risk for a patient with a history of complications (adjusting for the risk of second events) and covariance (e.g., patient age with duration of diabetes and/or history of complications, or ethnicity with baseline blood pressure) were also cited as major challenges, particularly when modeling without patient-level data. Different interpretations of end points were widely acknowledged as a reason that the model results did not match trial results in several cases. (For example, very few of the models captured revascularization as an end point, and differing methods were used to define and measure the occurrence of retinopathy end points.) These difficulties can be compounded when local treatment practice influences the end point. (For example, the clinical decision on when to perform revascularization can vary widely between regions and countries.)

Another limitation of most models highlighted during the discussion was their failure to report fatal and nonfatal events separately (even though this is an integral part of the modeling calculations going on in the background). This would seem to be an essential function of a model, and future improvements will need to address these shortcomings. Differing assumptions around the progression of risk factors over time in the modeling analyses were also raised as a barrier to matching the trial results. (For example, Hb A_{1c}, SBP, or serum lipid level changes over time.) It was also clear that different modelers had used different assumptions about other risk factors and this may have been a source of discrepancy in the results presented. As part of this discussion, the paucity of data on risk factor progression in type 2 diabetes was acknowledged. At present the only published formulae for risk factor progression are those from the UKPDS [14]. Although some modern trials have reported data on Hb A_{1c} progression (e.g., Fenofibrate Intervention and Event Lowering in Diabetes study provides data on Hb A_{1c} progression on modern therapy over 5 years), more work is needed in this area.

Methodological issues were also discussed at the Fifth Mount Hood Challenge, arising from the challenge simulations. In light of the data from the ACCORD trial, the following question was raised: What should the relationship between Hb A_{1c} levels and mortality be in diabetes models? Although some trial results suggest that the relationship could be U-shaped, more evidence is clearly needed to provide a definitive answer. This is an epidemiological puzzle that may become clearer in the years ahead. Another issue raised concerned the modeling of risk via changing risk factors as opposed to the direct treatment effects on end points (e.g., the gliclazide stroke effect). Most models currently rely on physiological risk factors (and patient characteristics) to estimate the risk of events. Although it would be advantageous to include specific treatment effects

Table 5 – Summary of validation results for a blood pressure–lowering intervention in the ACCORD trial.

Model		Primary end point	Death from any cause	CVD mortality	Fatal stroke	Major coronary disease event	Nonfatal MI	Any stroke	Fatal or nonfatal HF
ACCORD									
Intervention		1.9%	1.3%	0.5%		2.3%	1.1%	0.3%	0.7%
Control		2.1%	1.2%	0.5%		2.4%	1.3%	0.5%	0.8%
IMS CORE									
Intervention		1.0%	2.6%	0.4%	0.02%	0.9%	0.4%	0.2%	0.5%
Control		1.2%	2.7%	0.4%	0.03%	0.9%	0.5%	0.3%	0.6%
MICHIGAN									
Intervention		2.3%	1.7%	0.9%	0.2%	2.0%	1.0%	0.7%	
Control		2.8%	1.7%	1.0%	0.2%	2.4%	1.3%	0.9%	
ECHO-T2DM									
Intervention		2.2%	2.0%	0.9%	0.1%	2.0%	1.0%	0.4%	0.6%
Control		2.6%	2.0%	1.0%	0.2%	2.3%	1.1%	0.6%	0.7%
UKPDS OM									
Intervention		1.7%	2.7%	1.0%		1.8%		0.4%	0.6%
Control		1.9%	2.7%	1.0%		2.0%		0.5%	0.7%
UKPDS RE									
Intervention		1.9%							
Control		2.1%							
CDC-RTI									
Intervention		1.7%	2.9%	1.5%	0.4%			0.3%	
Control		1.9%	2.9%	1.6%	0.4%			0.4%	
CARDIFF									
Intervention		1.0%	1.2%	0.4%	0.0%		0.4%	0.3%	0.2%
Control		1.1%	1.1%	0.4%	0.1%		0.4%	0.4%	0.2%
EBMI									
Intervention			5.09%			2.52%		0.36%	2.43%
Control			5.15%			2.51%		0.61%	2.38%
ACCORD	Relative risk	0.88	1.07	1.06		0.94	0.87	0.59	0.94
IMS CORE	Relative risk	0.85	0.98	0.91	0.69	0.90	0.90	0.69	0.91
MICHIGAN		0.84	0.96	0.91	0.85	0.86	0.80	0.82	
ECHO-T2DM		0.85	0.98	0.90	0.85	0.87	0.86	0.69	0.85
UKPDS OM		0.88	0.98	1.02		0.92		0.77	0.89
UKPDS RE		0.87							
CDC-RTI		0.90	0.99	0.97	0.97			0.85	
CARDIFF		0.97	1.00	1.01	0.38		1.04	0.74	1.04
EBMI			0.99			1.00		0.60	1.02

Note. Relative risk represents the ratio of event risk in the intervention group versus the control group. ACCORD, Action to Control Cardiovascular Risk in Diabetes trial, a trial with median 4.7 years of follow-up in participants with type 2 diabetes at high risk for cardiovascular events, randomly assigned to intensive therapy (target SBP < 120 mm Hg) or standard therapy (target SBP < 140 mm Hg); CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; HF, heart failure; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; MICHIGAN, The Michigan Model for Diabetes; SBP, systolic blood pressure; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

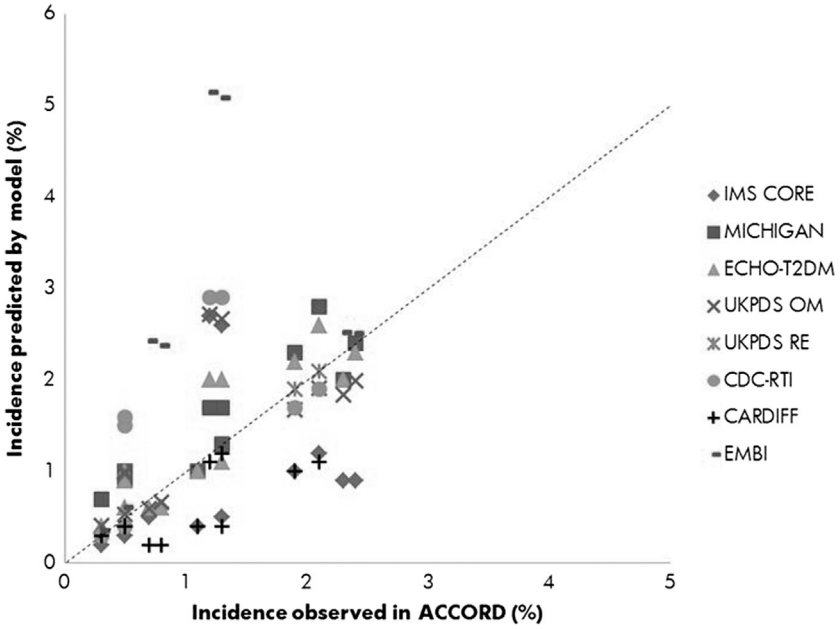

Fig. 4 – Summary of model validation results for the ACCORD blood pressure– lowering study. ACCORD, Action to Control Cardiovascular Risk in Diabetes trial, a trial with median 4.7 years of follow-up in participants with type 2 diabetes at high risk for cardiovascular events, randomly assigned to intensive therapy (target SBP <120 mm Hg) or standard therapy (target SBP <140 mm Hg); CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; HF, heart failure; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; MICHIGAN, The Michigan Model for Diabetes; SBP, systolic blood pressure; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

on end points in the models, suitable data are seldom available at the time of launch of new agents (when cost-effectiveness analyses are required to support reimbursement decision making).

Differences were acknowledged in the way models generate simulation cohorts. In some cases, simulation cohorts were directly generated from distributions based on the available published trial data. In others, an overall population was generated and then selected on the basis of clinical trial inclusion/exclusion criteria to create a simulation cohort. No consensus was reached on which of these approaches would be best, but it was agreed that access to patient-level data would improve the projections made by most of the models presented at the Fifth Mount Hood Challenge. Although access to patient-level data from trials is frequently highly restricted, and patient-level simulation places more demands on computing resources, such data do allow covariance between different risk factors to be captured, and this may have important implications for the accuracy of simulations. To this end, electronic medical registry data may prove to be a valuable resource for estimating covariance matrices. They may also contain a much more heterogeneous set of patients who are more representative of the general population than are trials with restrictive recruitment criteria. Such data sets, however, may also have other selection biases and frequently lack rigorous clinical adjudication of end-point events in comparison with clinical trials. The only models presented at the Mount Hood Meeting that captured the influence of covariance were the UKPDS Outcomes Model and the ECHO-T2DM Model, both of which use a covariance matrix developed from patient-level data. The influence of this matrix on model outcomes has not yet been fully investigated. It may, however, have been one of the reasons why the UKPDS Outcomes Model and the other models at the meeting that rely on individual elements of the same UKPDS regression equations to estimate risk produced different results in the validation exercises.

The influence of ethnic characteristics on the risk of complications in patients with type 2 diabetes was also raised as a point of methodology. While it would be optimal to factor this fully into the modeling analyses, ethnic group information is often poorly recorded, confounded with socioeconomic status, and/or subject to high uncertainty because of small numbers.

Although the Fifth Mount Hood Challenge Meeting centered primarily on validation exercises as a vehicle to compare and contrast modeling methodologies in different groups, it was clear from the meeting that there is no clear consensus on precisely what model validation means. Appropriate statistical approaches should be defined to assess correlation between model and clinical trial outcomes, and limits could be predefined for model accuracy and precision. A consensus group with appropriate statistical expertise may offer the best opportunity to resolve this long-standing issue.

The Fifth Mount Hood Challenge Meeting included a session on dealing with statistical uncertainty in simulation models of type 2 diabetes. Prof. Andrew Briggs highlighted the importance and many of the challenges in terms of dealing with statistical uncertainty in complex disease models. The importance of capturing parameter uncertainty (and dealing with parameter estimation) was emphasized because it affects decision uncertainty. Structural uncertainty in models, an aspect that is often overlooked, is being tackled in part by the modeling comparison

Table 6 – Summary of validation results for a glucose-lowering intervention in the ACCORD trial.

Model		Duration (y)	Primary end point	All-cause mortality	CVD mortality	Nonfatal MI	Nonfatal stroke
ACCORD							
Intervention		3.5	6.9%	5.0%	1.7%	3.6%	1.3%
Control			7.2%	4.0%	1.3%	4.6%	1.2%
IMS CORE							
Intervention		4		10.1%	1.5%	1.9%	0.9%
Control				10.2%	1.9%	2.2%	1.0%
ECHO-T2DM							
Intervention		3.5	8.1%	9.4%	3.6%	3.5%	1.3%
Control			9.0%	10.0%	4.1%	3.8%	1.5%
UKPDS OM							
Intervention		3.5	6.7%	10.6%	4.6%		
Control			7.4%	11.1%	4.7%		
UKPDS RE							
Intervention		3.5	6.3%				
Control			7.1%				
CDC-RTI							
Intervention		4		11.2%	5.9%		
Control				11.3%	6.1%		
CARDIFF							
Intervention		4		4.8%	1.1%	1.6%	0.8%
Control				4.8%	1.3%	1.7%	0.9%
EBMI							
Intervention		3.5		17.3%		3.4%	1.6%
Control				12.4%		4.5%	1.8%
ACCORD	Relative risk	3.5	0.96	1.25	1.31	0.78	1.08
IMS CORE		4		0.99	0.79	0.87	0.83
ECHO-T2DM		3.5	0.90	0.94	0.88	0.91	0.88
UKPDS OM		3.5	0.90	0.96	0.97		
UKPDS RE		3.5	0.90				
CDC-RTI	Relative risk	4		0.99	0.97		
CARDIFF		4		1.01	0.84	0.96	0.90
EBMI		3.5		1.39		0.74	0.88

Notes. The Michigan Model did not report outcomes for the (optional) glucose- lowering intervention analysis from the ACCORD trial. Relative risk represents the ratio of event risk in the intervention group versus the control group. ACCORD, Action to Control Cardiovascular Risk in Diabetes trial, a trial with median 3.5 years of follow-up in participants with type 2 diabetes at high risk for cardiovascular events, randomly assigned to receive intensive therapy (target Hb A_{1c} value below 6.0%) or standard therapy (targeting Hb A_{1c} value of 7.0%–7.9%); CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; Hb A_{1c}, hemoglobin A_{1c}; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model; UKPDS RE, the United Kingdom Prospective Diabetes Study Risk Engine.

at the Mount Hood meetings. Validation of simulation models has an important role to play in improving modeling efforts and, similarly, meetings such as the Mount Hood challenges offer a unique environment to further this cause. The Fifth Mount Hood Challenge modelers have performed an additional set of challenge simulations designed to investigate statistical uncertainty within the individual models (as well as aspects of structural uncertainty), and this analysis will be the subject of a future collaborative article.

Data from the UKPDS have revolutionized the modeling of type 2 diabetes. Although UKPDS patients continued to be followed up until 2007, there is clearly a need for additional patient-level data sets to better understand treatment innovations, novel risk factors, and different target populations. As the diabetes epidemic continues to grow, particularly in the developing world, an already complex environment for modelers will give rise to even more challenges. It could be that the future for diabetes modeling will rely on the development of country-specific models or at least country-specific/population-specific risk estimates. Socioeconomic status and the role of molecular genetic testing may have an important role to play in future. An alternative scenario could see collaborative methods, such as those in the field of climate change, where recommendations are made on the basis of averages from the results of more than 20 different models [25]. Regardless of these future avenues of research, meetings such as the Mount Hood challenges will have an important role to play as modelers seek to continually improve on the performance of their diabetes models to better meet the needs of health care decision makers around the world. Modelers in other disease areas might wish to consider whether adopting a similar process in their area would have similar benefits in identifying problems and accelerating improvements.

Acknowledgments

The Mount Hood 5 Modeling Group is grateful to our corporate sponsors for funding to support the organization of the meeting

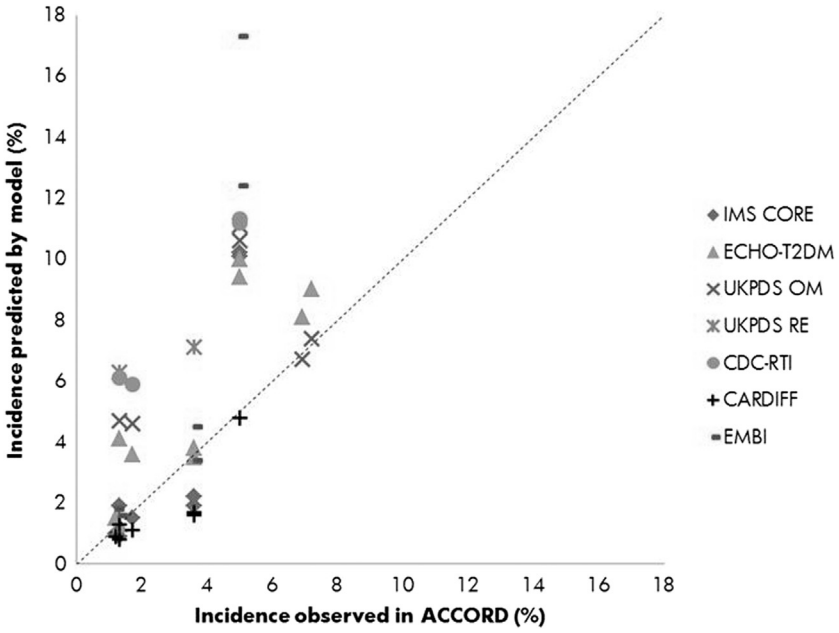

Fig. 5 – Summary of model validation results for the ACCORD glucose-lowering study. The Michigan Model did not report outcomes for the (optional) glucose-lowering intervention analysis from the ACCORD trial. ACCORD, Action to Control Cardiovascular Risk in Diabetes trial, a trial with median 3.5 years of follow-up in participants with type 2 diabetes at high risk for cardiovascular events, randomly assigned to receive intensive therapy (target Hb A_{1c} value below 6.0%) or standard therapy (targeting Hb A_{1c} value of 7.0%–7.9%). CARDIFF, Cardiff Research Consortium Model; CDC-RTI, the Centers for Disease Control and Prevention-RTI Diabetes Cost-effectiveness Model; CVD, cardiovascular disease; EBMI, the Evidence-Based Medicine Integrator Simulator; ECHO-T2DM, The Swedish Institute of Health Economics model titled Economics and Health Outcomes in Type 2 Diabetes Mellitus Model; IMS CORE, IMS CORE Diabetes Model; MI, myocardial infarction; UKPDS OM, the United Kingdom Prospective Diabetes Study Outcomes Model.

and the publication of this manuscript: Novo Nordisk A/S (Copenhagen, Denmark) and Janssen Pharmaceutica NV (Beerse, Belgium).

The Mount Hood 5 Modeling Group authors that have contributed to this paper include: Andrew Palmer, Menzies Research Institute Tasmania, University of Tasmania, Australia; Philip Clarke, School of Public Health, University of Sydney, Australia; Alastair Gray (UKPDS Outcomes Model), Health Economics Research Centre, University of Oxford, UK; Jose Leal (UKPDS Outcomes Model), Health Economics Research Centre, University of Oxford, UK; Adam Lloyd (IMS CORE Diabetes Model), IMS Health, London UK; David Grant (IMS CORE Diabetes Model), IMS Health, London UK; James Palmer (IMS CORE Diabetes Model), IMS Health, Basel, Switzerland; Volker Foos (IMS CORE Diabetes Model), IMS Health, Basel, Switzerland; Mark Lamotte (IMS CORE Diabetes Model), IMS Health, Brussels, Belgium; William Hermann (Michigan Model), Michigan Diabetes Research and Training Center, University of Michigan, MI, USA; Jacob Barhak (Michigan Model), University of Michigan, MI, USA; Michael Willis (ECHO/IHE Model), The Swedish Institute of Health Economics, Lund, Sweden; Ruth Coleman (UKPDS Risk Engine), UKPDS Group Diabetes Trials Unit, University of Oxford, UK; Ping Zhang (CDC/RTI model), Centers for Disease Control and Prevention, Atlanta, GA, USA; Phil McEwan (Cardiff Model), Cardiff Research Consortium, Cardiff, UK / Centre for Health Economics, University of Swansea, UK; Jonathan Betz Brown (EBMI Simulator), Portland, OR, USA; Ulf Gerdtham, University of Lund, Sweden; Elbert Huang, University of Chicago, IL, USA; Andrew Briggs, University of Glasgow, UK; Katarina Steen Carlsson, The Swedish Institute of Health Economics, Lund, Sweden; and William Valentine, Ossian Health Economics and Communication, Basel, Switzerland.

Source of financial support: The organization of the Mount Hood 5 Meeting was supported by an unrestricted grant from Novo Nordisk A/S (Copenhagen, Denmark) and Janssen Pharmaceutica NV (Beerse, Belgium).

REFERENCES

[1] Brown JB, Palmer AJ, Bisgaard P, et al. The Mt. Hood challenge: cross-testing two diabetes simulation models. Diabetes Res Clin Pract 2000;50(Suppl. 3):S57–64.
[2] International Diabetes Federation Diabetes Atlas Fourth Edition. Available from: http://www.diabetesatlas.org/content/diabetes-and-impaired-glucose-tolerance. [Accessed May 23, 2011].
[3] Mount Hood 4 Modeling Group. Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting. Diabetes Care 2007;30:1638–46.
[4] Mount Hood Modeling Group Web site. Available from: https://sites.google.com/site/mounthoodchallenge/. [Accessed January 17, 2013].
[5] Knopp RH, d'Emden M, Smilde JG, Pocock SJ. Efficacy and safety of atorvastatin in the prevention of cardiovascular end points in subjects with type 2 diabetes: the Atorvastatin Study for Prevention of Coronary Heart Disease Endpoints in non-insulin-dependent diabetes mellitus (ASPEN). Diabetes Care 2006;29:1478–85.
[6] ADVANCE Collaborative Group, Patel A, MacMahon S, Chalmers J, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. N Engl J Med 2008;358:2560–72.
[7] ACCORD Study Group, Cushman WC, Evans GW, Byington RP, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med 2010;362:1575–85.
[8] ACCORD Study Group, Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008;358:2545–59.
[9] Palmer AJ, Roze S, Valentine WJ, et al. The CORE Diabetes Model: projecting long-term clinical outcomes, costs and cost-effectiveness of

interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision making. Curr Med Res Opin 2004;20(Suppl. 1):S5–26.
[10] Palmer AJ, Roze S, Valentine W, et al. Validation of the CORE Diabetes Model against epidemiological and clinical studies. Curr Med Res Opin 2004;20(Suppl. 1):S27–40.
[11] Zhou H, Isaman DJM, Messinger S, et al. A computer simulation model of diabetes progression, quality of life and cost. Diabetes Care 2005;28:2835–63.
[12] Barhak J, Isaman DJM, Ye W, Lee D. Chronic disease modeling and simulation software. J Biomed Inform 2010;43:791–9.
[13] Asseburg C, Willis M, Persson U, et al. Validation of the IHE/J&J economic simulations model of type 2 diabetes mellitus (T2DM). Presented at the 14th Annual ISPOR International Meeting. May 18, 2009.
[14] Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS 68). Diabetologia 2004;47:1747–59.
[15] Stevens RJ, Kothari V, Adler AI, et al. The UKPDS risk engine: a model for the risk of coronary heart disease in type II diabetes (UKPDS 56). Clin Sci (Lond) 2001;101:671–9.
[16] Kothari V, Stevens RJ, Adler AI, et al. UKPDS 60: risk of stroke in type 2 diabetes estimated by the UK Prospective Diabetes Study risk engine. Stroke 2002;33:1776–81.
[17] Stevens RJ, Coleman RL, Adler AI, et al. Risk factors for myocardial infarction case fatality and stroke case fatality in type 2 diabetes: UKPDS 66. Diabetes Care 2004;27:201–7.
[18] CDC Diabetes Cost-effectiveness Group. Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes. JAMA 2002;287:2542–51.
[19] Hoerger T.J., Segel J.E., Zhang P., Sorensen S.W. Validation of the CDC-RTI Diabetes Cost-Effectiveness Model. RTI Press Methods Report, 2009. Available from: http://www.rti.org/pubs/mr-0013-0909-hoerger.pdf. [Accessed February 2, 2012].
[20] CDC-RTI Diabetes Cost-Effectiveness Model description. Available from: http://care.diabetesjournals.org/content/suppl/2010/06/14/dc10-0554.DC1/dc100554onlineappendix.pdf. [Accessed February 2, 2012].
[21] McEwan P, Bergenheim K, Yuan Y, et al. Assessing the relationship between computational speed and precision: a case study comparing an interpreted versus compiled programming language using a stochastic simulation model in diabetes care. Pharmacoeconomics 2010;28:665–74.
[22] McEwan P, Evans M, Kan H, Bergenheim K. Understanding the inter-relationship between improved glycaemic control, hypoglycaemia and weight change within a long-term economic model. Diabetes Obes Metab 2010;12:431–6.
[23] McEwan P, Peters JR, Bergenheim K, Currie CJ. Evaluation of the costs and outcomes from changes in risk factors in type 2 diabetes using the Cardiff stochastic simulation cost-utility model (DiabForecaster). Curr Med Res Opin 2006;22:121–9.
[24] Blum S, Vardi M, Brown JB, et al. Vitamin E reduces cardiovascular disease in individuals with diabetes mellitus and the haptoglobin 2-2 genotype. Pharmacogenomics 2010;11:675–84.
[25] Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report Climate Change 2007: Working Group I: The Physical Science Basis. Available from: http://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch10.html. [Accessed March 9, 2012].